Davide Belli
11887532
davidebelli95@gmail.com

Homework Assignment 7
Machine Learning 3, 17/18

2018-05-21

In this Homework, I discussed possible interpretations with Gabriele Cesa.

1 Problem 1

1.1 Question 1

Note the typos in the last derivations for γ and Σ: the matrices in the third terms of the rispective numerators should be transposed, as: $(A^{new})^T$ and $(C^{new})^T$

1) $A^{new} = \underset{A}{\arg\max}\, Q(\theta, \theta^{old})$

$$0 \overset{!}{=} \frac{\partial}{\partial A}\left(-\frac{N-1}{2}\ln|\Gamma| - \mathbb{E}_{z|\theta^{old}}\left[\frac{1}{2}\sum_{n=2}^{N}(z_n - A z_{n-1})^T \Gamma^{-1}(z_n - A z_{n-1})\right] + Cst\right)$$

$$0 = \sum_{n=2}^{N}\mathbb{E}_z\left[\cancel{\Gamma^{-1}}(-A z_{n-1} + z_n) z_{n-1}^T\right]$$

$$A \sum_{n=2}^{N}\mathbb{E}_z\left[z_{n-1} z_{n-1}^T\right] = \sum_{n=2}^{N}\mathbb{E}_z\left[z_n z_{n-1}^T\right]$$

$$A^{new} = \left(\sum_{n=2}^{N}\mathbb{E}_z\left[z_n z_{n-1}^T\right]\right)\left(\sum_{n=2}^{N}\mathbb{E}_z\left[z_{n-1} z_{n-1}^T\right]\right)^{-1}$$

$$\Gamma^{new} = \underset{\Gamma}{\arg\max}\, Q(\theta, \theta^{old})$$

$$0 \overset{!}{=} \frac{\partial}{\partial \Gamma} Q(\theta, \theta^{old})$$

$$\cancel{\Gamma^{-1}}(N-1) = \cancel{\Gamma^{-1}}\sum_{n=2}^{N}\mathbb{E}_z\left[(z_n - A^{new} z_{n-1})(z_n - A^{new} z_{n-1})^T\right]\Gamma^{-1}$$

$$\Gamma = \frac{\sum_{n=2}^{N}\mathbb{E}_z\left[(z_n - A^{new} z_{n-1})(z_n - A^{new} z_{n-1})^T\right]}{N-1}$$

$$\Gamma = \frac{\sum_{n=2}^{N}\left(\mathbb{E}_z[z_n z_n^T] - A^{new}\mathbb{E}_z[z_{n-1} z_n^T] - \mathbb{E}_z[z_n z_{n-1}^T] A^{new} + A^{new}\mathbb{E}_z[z_{n-1} z_{n-1}^T](A^{new})^T\right)}{N-1}$$

1.2 Question 2

2) Similarly to the previous derivations, we find:

$$C^{new} = \frac{\partial}{\partial C} Q(\theta, \theta^{old})$$

and

$$C^{new} = \left(\sum_{n=1}^{N} \mathbb{E}_z[x_n z_n^T]\right)\left(\sum_{n=1}^{N} \mathbb{E}_z[z_n z_n^T]\right)^{-1}$$

$$= \left(\sum_{n=1}^{N} x_n \mathbb{E}_z[z_n^T]\right)\left(\sum_{n=1}^{N} \mathbb{E}_z[z_n z_n^T]\right)^{-1}$$

$$\Sigma^{new} = \frac{\partial}{\partial \Sigma} Q(\theta, \theta^{old})$$

$$\Sigma^{new} = \frac{\sum_{n=1}^{N} \mathbb{E}_z[(x_n - C^{new} z_n)(x_n - C^{new} z_n)^T]}{N}$$

$$= \frac{\sum_{n=1}^{N} \left(x_n x_n^T - C^{new}\mathbb{E}_z[z_n]x_n^T - x_n \mathbb{E}_z[z_n^T] C^{new} + C^{new}\mathbb{E}_z[z_n z_n^T](C^{new})^T\right)}{N}$$

2 Problem 2

Note that A, B, C are just names for the different causal model configurations.

	A	B	C
a)	$X \rightarrow Y$	$X \leftarrow Y$	$X \quad Y$
b)	$p(x,y) = p(x)p(y\|x)$	$p(x,y) = p(y)p(x\|y)$	$p(x,y) = p(x)p(y)$
c)	$p(y\|x) = \frac{p(x,y)}{p(x)} = \frac{p(y\|x)\cancel{p(x)}}{\cancel{p(x)}} = p(y\|x)$	$p(y\|x) = \frac{p(x,y)}{p(x)} = \frac{p(y)p(x\|y)}{p(x)}$	$p(y\|x) = \frac{p(x,y)}{p(x)} = \frac{\cancel{p(x)}p(y)}{\cancel{p(x)}} = p(y)$

d) By applying do(x), the causal graph reduces to $G_{\overline{X}}$:

A	B	C
$X \rightarrow Y$	$X \quad Y$	$X \quad Y$
$p(y\|do(x)) = p(y\|x)$	$p(y\|do(x)) = p(y)$	$p(y\|do(x)) = p(y)$

e) p(cancer|smokes) is the conditional probability for someone to have cancer if we know that he smokes. Note that the conditioning over smoking is an observation on the available data. As a consequence, there may be some latent cause for cancer and smokes that explain possible correlations.
On the other way, p(cancer|do(smokes)) represents the probability of a person to have cancer if we force it to smoke. This means that we remove every possible external cause that could be the reason for him to smoke and that could possible also influence the probability of having cancer.

The main difference, as we explained, is in the fact that with the do operator we influence the system externally, removing all possible causes of a variable by manually tuning its value. Conditioning is a simple observation over the value of a variable.

3 Problem 3

3.1 Question 1

(Recovery | Drug) rate = 50%
(Recovery | No Drug) rate = 40%

I would advice a new patient to take the drug since it seems to increase the recovery rate given the available information.

3.2 Question 2

(Recovery | Drug, Male) rate = 60%
(Recovery | No Drug, Male) rate = 70%

(Recovery | Drug, Female) rate = 20%
(Recovery | No Drug, Female) rate = 30%

I would advice both patients to not take the drug since, given the study case results divided by gender, we see that the recovery rate for both sexes decreases when assuming the drug.

3.3 Question 3

I would advice the patient to not take the drug. We see that the men have higher recovery rate independently on the use of the drug with respect to woman, and the number of men and women taking and not taking the drug in the experiment was unbalanced. As a consequence, the initial result is biased by an unbalance in the experiment, and external cause which is the difference in recovery rates per sex (note that this in an assumption driven by the available data). If we compute the probabilities of recovery balancing the classes (e.g. dividing/multiplying by three the values for men or women), we find that the Recovery rate overall is :

(Recovery | Drug) rate = 40%
(Recovery | No Drug) rate = 50%

For this reason, I would contradict my earlier advice in light of the new observations on the test subjects.

3.4 Question 4

Applying the backdoor criterion, we see that M is admissible for adjustment to find the causal effect of D on R, since the backpath is blocked by the non-collider M. The probabilty results in:

$p(R|do(D)) = \int p(r|d,m)p(m)dm$
In this case, I would advice the the patient to not take the drug, for the same reasons discussed previously.

3.5 Question 5

In this case, using the backdoor criterion we note that there are no back paths starting with an incoming edge in D. For this reason, the empty set is admissible for adjustment to find the causal effect of D on R, namely D is the cause of R (possibly via M), and the probabilities simplifies to:

$p(R|do(D)) = p(R|D)$
In this case, I would suggest to take the drug.

3.6 Question 6

An interpretation for the variables could be as follows

1. L1 is a latent variable describing where the subject lives (in wealthier countries access to drugs is easier), which, in anonymous studies, we assume not to be shared.
2. L2 is some gene in the subject that also correlates with the effecacy of some medical compounds on the body.
3. M can be, for example, how much the subject is tanned. This depends both on where they live and also on their genetycs. Also, we assume the tanning do not influence the subject in taking the drugs or on recovery rates.

In this case, M cannot be used to express the causal effect of D on R, since it is a collider belonging on the conditioning set when applying backdoor criterion to paths from R to D. We notice that every subset of {L2, L1, M} containing one of the two random variables contains a blocking non-collider which could explain the causal effect. However, those latent variables are not measurable and so we cannot practically use them to run experiments. Finally, we notice that the empty set is admissible for adjustment of the causall effect, since the variable M is a collider not belonging to the conditioning set. As a result, we once again express the probability as:
$p(R|do(D)) = p(R|D)$
Under this interpretation, M is uninfluent on the effects of the drug, and once again I would suggest to take the drug.

4 Problem 4

4.1 Question 1

$$p(R,W,S) = p(R)p(S)p(W|R,S)$$
$$p(R) = p(E_R)$$

$$
\begin{aligned}
p(S) &= p(E_S) \\
p(W) &= \sum_{R,S} p(R,W,S) \\
&= \sum_{R,S} p(R)p(S)p(W|R,S) \\
p(W=0) &= p(R=0)p(S=0) = 0.3 * 0.6 = 0.18 \\
p(W=1) &= 1-(1-p(R=1))(1-p(S=1)) \\
&= 1-p(W=0) = 0.82
\end{aligned}
$$

4.2 Question 2

$$
\begin{aligned}
p(R=1|W=1) &= \frac{p(W=1|R=1)p(R=1)}{p(W=1)} \\
&= \frac{1*0.7}{0.82} \approx 0.8537
\end{aligned}
$$

4.3 Question 3

No, it just means that the two variables are not independent. When correlation occurs between two variables, it can be that one causes the other, or that some other variable is a common direct or indirect cause of them.

4.4 Question 4

All the three variables in the graph are independent since we remove from the causal graph all incoming edges for the intervened variable. Thre resulting probabilities are:

$$
\begin{aligned}
p(R,W,S) &= p(R)p(S)p(W) \\
p(R) &= p(E_R) \\
p(S) &= p(E_S) \\
p(W=0) &= 0 \\
p(W=1) &= 1
\end{aligned}
$$

4.5 Question 5

Using do-calculus:

$$
\begin{aligned}
(R \not\perp\!\!\!\perp W)_{\mathcal{G}_{\underline{W}}} &\implies p(R|do(W=w)) \neq p(R|W=w) \\
(R \perp\!\!\!\perp W)_{\mathcal{G}_{\overline{W}}} &\implies p(R|do(W=w)) = p(R)
\end{aligned}
$$

4.6 Question 6

The causal graph remains the same, except for a change in the marginal over S:

$$
\begin{aligned}
p(R,W,S) &= p(R)p(S)p(W|R,S) \\
p(R) &= p(E_R) \\
p(S=1) &= \mathbb{I}(s=1) \\
p(S=0) &= \mathbb{I}(s=0) \\
p(W) &= \sum_{R,S} p(R,W,S) \\
&= \sum_{R,S} p(R)p(S)p(W|R,S) \\
p(W=0) &= p(R=0)p(S=0) = 0.3 * \mathbb{I}(s=0) \\
p(W=1) &= 1-(1-p(R=1))(1-p(S=1)) \\
&= 1-p(W=0) = 1-0.3 * \mathbb{I}(s=0)
\end{aligned}
$$

4.7 Question 7

Using do-calculus:

$$
\begin{aligned}
(S \perp\!\!\!\perp W)_{\mathcal{G}_{\underline{S}}} &\implies p(W|do(S=s)) = p(W|S=s) \\
(S \not\!\perp\!\!\!\perp W)_{\mathcal{G}_{\overline{S}}} &\implies p(W|do(S=s)) \neq p(W)
\end{aligned}
$$

5 Problem 5

5.1 Question 1

Note that in the second step we factorize the numerator using the Truncated Factorization theorem. The denominator, as explained in the next task, is simpliefed using Ignoring Actions property of do-calculus.

$$
\begin{aligned}
p(Y|do(X), \boldsymbol{X}_{pa(X)}) &= \frac{p(Y, \boldsymbol{X}_{pa(X)}|do(X))}{p(\boldsymbol{X}_{pa(X)}|do(X)} \\
&= \frac{p(Y|X, \boldsymbol{X}_{pa(X)})p(\boldsymbol{X}_{pa(X)})}{p(\boldsymbol{X}_{pa(X)})} \\
&= p(Y|X, \boldsymbol{X}_{pa(X)})
\end{aligned}
$$

5.2 Question 2

in the second-last step, we respectively apply to the first and second term the result from the previous task and Ignoring Actions property, where:
$(\boldsymbol{X}_{pa(X)} \perp\!\!\!\perp X)_{\mathcal{G}_{\overline{X}}} \implies p(\boldsymbol{X}_{pa(X)}|do(X)) = p(\boldsymbol{X}_{pa(X)})$

$$
\begin{aligned}
p(Y|do(X)) &= \int p(Y, \boldsymbol{X}_{pa(X)}|do(X)) d\boldsymbol{X}_{pa(X)} \\
&= \int p(Y|do(X), \boldsymbol{X}_{pa(X)}) p(\boldsymbol{X}_{pa(X)}|do(X)) d\boldsymbol{X}_{pa(X)} \\
&= \int p(Y|X, \boldsymbol{X}_{pa(X)}) p(\boldsymbol{X}_{pa(X)}) d\boldsymbol{X}_{pa(X)}
\end{aligned}
$$